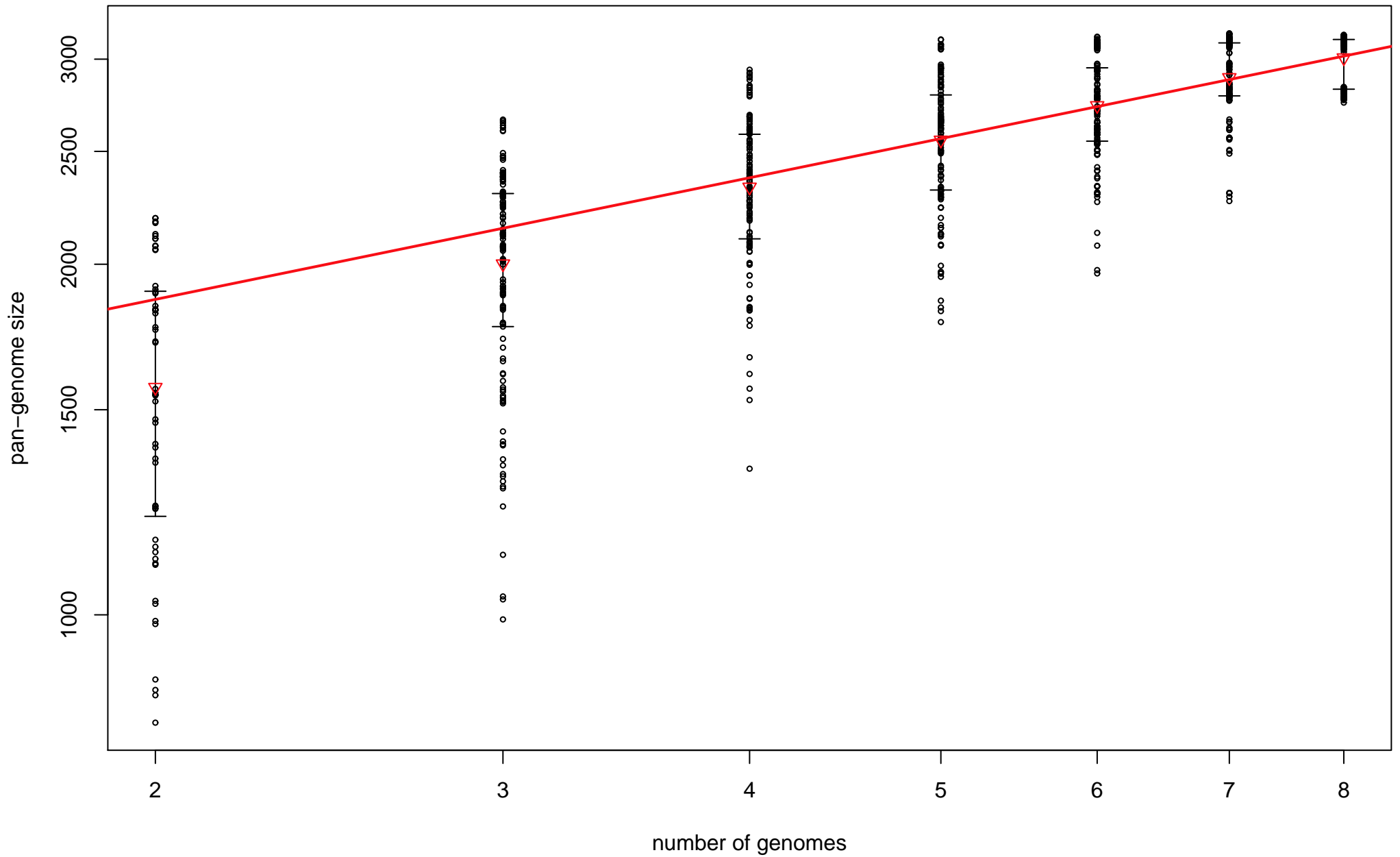
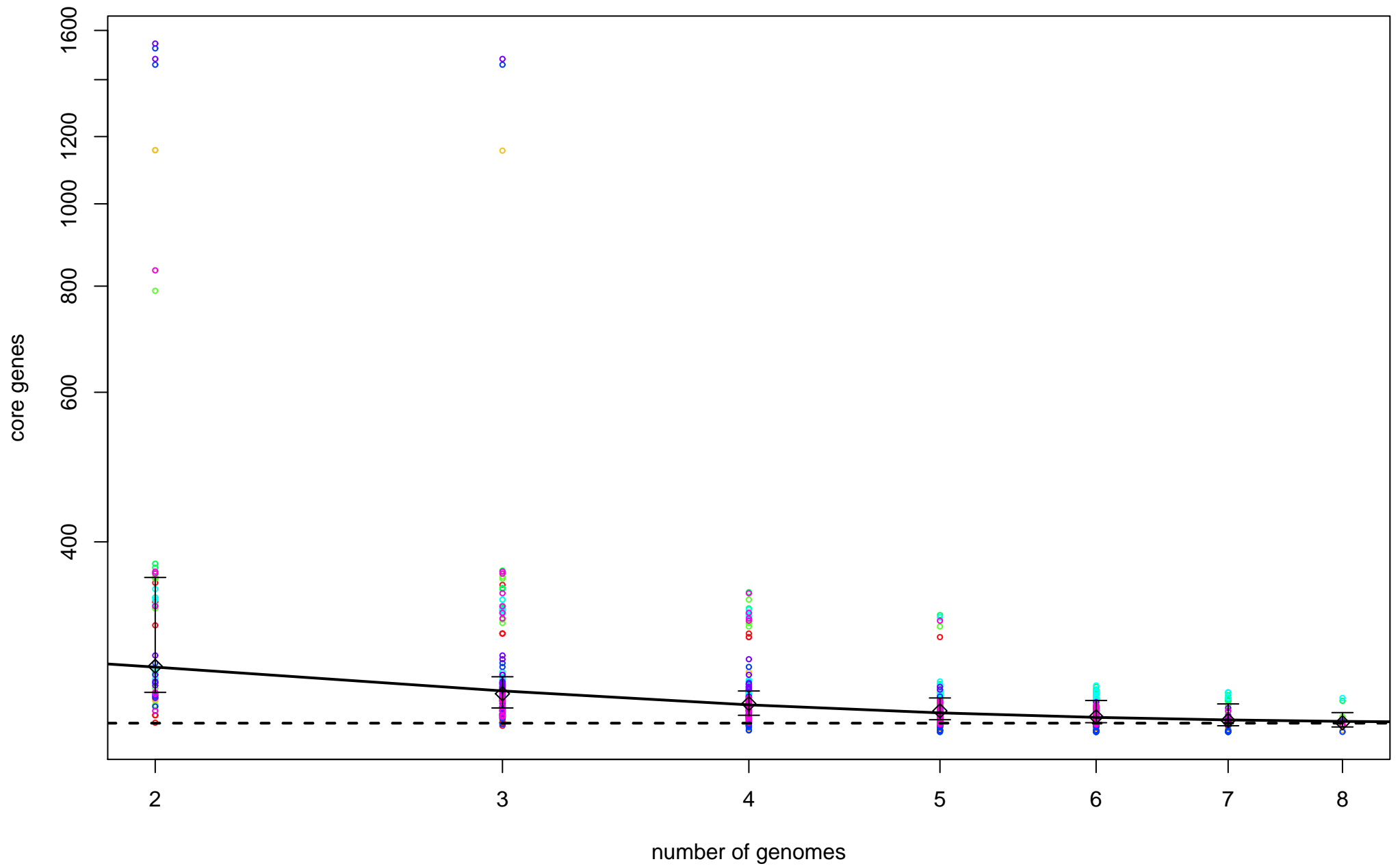


pan-genome power law log axis 5 or more genomes



$y = 1466.927 \pm 35.581x^{(0.3469 \pm 0.0129)}$

core genes exponential log axis



$y = 244.6938 \pm 1.4068 + 129.3912 \pm 16.9289e^{(-x/(1.7092 \pm 0.2018))}$

Figure S2. Pan and core-genome plots of the hemoplasmas. Plots were generated as previously described [24]. The number of genes is represented as a function of the number of sequenced genomes (x-axes: number of genomes; y-axes: number of genes). Genomes are sequentially sampled in all combinations until the last genome is analyzed. Error bars represent the 1st and 3rd quartile of these samples, and triangles (pan-genome plot) and diamonds (core-genome plot) represent the medians. The power law function is fit to all medians. The following genomes were used: *M. haemocanis* str. Illinois (PRJNA82367), *M. suis* str. Illinois (PRJNA61897), *M. haemofelis* strain Ohio2 (PRJNA162029), *M. wenyonii* str. Massachusetts (PRJNA168067), ‘*Candidatus M. haemolamae*’ str. Purdue (PRJNA68151), ‘*Candidatus M. haemominutum*’ str. Birmingham 1 (PRJNA76933), *M. suis* strain KI3806 (PRJNA63665), and *M. haemofelis* strain Langford1 (PRJNA62461).